International Journal of Pure and Applied Mathematics
Volume 118 No. 24 2018
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

DETECTION OF MALICIOUS APPLICATION ON LINUX POWERED DEVICES USING SVM CLASSIFIERS

Mr.Rahul Y. Pawar[1], Dr.C.Mahesh[2],
[1] Department of Computer Engineering,
[2]Department of Information Technology,
[1]D.Y.Patil College of Engineering, [2]Veltech University
[1]Pune, India
[2]Chennai, India.
rahulpawar8087@gmail.com
chimahesh@gmail.com

May 26, 2018

Abstract

In Earlier days the challenges for security in personal computers and desktop computers were having huge scope and now a days they are becoming major issue in Smartphone security. As a Practice, Well recognized Security Systems for desktop are considered and evaluation done whether they are applicable to mobile devices or not. However, some of the static systems for desktop like antivirus software are not suitable for Mobile devices as their CPU Consumption is very high also they consume too much memory and the result is rapid draining of the Battery power. In static malware detection systems like antivirus detection capabilities and accuracy depends on the Database which contains of a recent malware signatures, therefore if an attacker spreads previously un-encountered malware the antivirus users are not protected. If we considered as per the

response time of static systems like antivirus it can be between several hours to several days as the process has three steps identifying the new virus, generating a signature, and updating the database, so attackers have high chance of attacking. Many malware instances are also targeting a specific and relatively small features in mobile devices like extracting confidential information and tracking mobile devices location, therefore they are slow movers and difficult to identify.

***Key Words*:**Malware Detection, Linux powered devices, Android, Machine learning classifiers, SVM algorithm.

1 Introduction

As per Market survey Android Operating System has most shares in the total market-share which is more than 68%, leaving its competitors Samsung, Windows, I phone, Nokia, Blackberry and Java far behind if compared. The use of smart phones started in the last decade, but the launch of Android and IOS has changed the market scenario by creating an enormous attraction all over the world among customers and application developers. Smartphones have become Pervasive due to availability of connections over the network, such as CDMA, GSM, EDGE, Wi-Fi, GPS, LTE, and Bluetooth. As per Market share report of year 2017 it is observed that there is rapid increase of 34.3% in Smartphone sales from 2016. The comparison of total sales in year 2017 is shown in the Figure 1.1. It is also analyzed that there is tremendous increase in number of sells for mobile devices having Android operating systems than other competitors whose sales are going down. Always connectivity with internet and important personal information such as bank transactions, social network access, contacts, messages, and browsing history have given attackers good chance for malware development. When OS platform is considered Android in Particular is in hit list of the attackers due to its popularity. In this way is can be simply observed that there is increase in malicious applications has challenged anti-malware industry to design and develop effective methods for dynamic detection of malwares in available constraints. Majority of antivirus software's still having static signature based detection due to simplicity and efficiency in terms of

implementation.

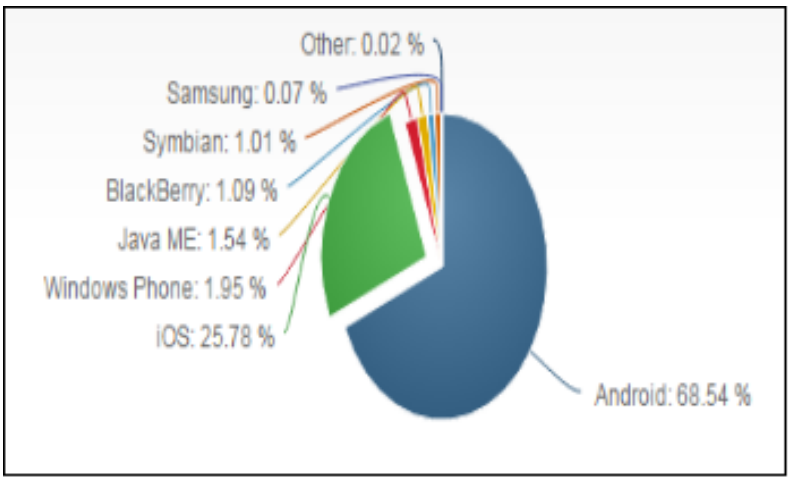

Figure 1 Market share survey 2017

Signature based methods can be easily penetrated with code obfuscation, and even with small modification in code. Requirement of a new signature for every new malware that is the reason why an anti-virus user has to update its signature database regularly. Due to some limitations like low processing capability and low battery power on a mobile device, cloud-based solutions for malware detection came into picture. Due to rapid increase in malwares and their clones, there is a requirement to design automatic and dynamic detection system with high accuracy and efficiency.

2 PRILIMINARY

Now a day's Internet of Things is on boom and internet is very dangerous for your Mobile phones in terms of files, websites and emails. Often attackers do active attacks by attacking trusted games, or WebPages. Entities like malware, Viruses, worms, Trojan horses spyware, adware, and gray wares can penetrate your Smartphones security by blocking your internet connection, system crash, and using your personal information. Prevention is the best way to

protect the device from such attacks, and it can be done by creating the awareness regarding the activities done by the threats before they can attack on device. Malware is a broad term for malicious software that can show the characteristics of viruses, Trojan horses, worms, adware, gray ware, and others. Malware is created to execute parallel with normal system operation, usually giving attackers a slide edge to gain access to your mobile device and steal important information.

A. Malware Detection in Mobile Devices

Due to increasing storage capabilities and high computing power attackers emerge against mobile devices with new threats. The solution is to design Server based Malware detection models for mobile devices. As two solutions are possible static as well as dynamic detection techniques for malware detection. The Proposed system shows that host-based approaches with dynamic detection algorithms are required, since static algorithms are not sufficient to prevent the system from the upcoming malwares. Some of the data types can be used to design malware detection models, and finally solution can be proposed with dynamic framework for malware detection.

Large number of publications is available in the field of malware detection on mobile phones in the past years where no such focus is on dynamic detection using system calls. In first set of approaches battery power based frameworks were introduced in where the dependency was on the high life of the battery power. In second set of approaches Dynamic and behavior-based approaches came into picture, but the problem was such systems suffer from an additional overhead of computation so requirement of an External server was the issue. Irrespective of these publications, we can use some of system level and kernel-level features. So as a solution a light-weight dynamic detection framework can be designed for investigation of smart phones. Andromaly is a framework which is combination of anomaly and android. The proposed framework includes a Server based system for Malware Detection that will monitor continuously various features and system calls obtained from the Smartphone and can be sent to Machine Learning algorithms to classify the collected data as legitimate or malicious application. Evaluation done with several combinations of machine learning algorithms as well as various feature selection algorithms for selection of features with

high information gain. The overall system yields the high performance and accuracy even in new malware detection on Android smart phones.

3 PROBLEM STATEMENT

Mobile devices and recently new smart phones have continuously are in evolution phase from simple mobile devices into smarter and efficient minicomputers which can be used as internet of things even web of things when connected to a wide networks. Smart phones are designed as open source, and easy to program, so they are susceptible to various attacks done by various malwares, threats, viruses, and worms, all of which are also present for computer platforms. Such devices allow users open access to use the Internet connection over the network. Even for SMSs, MMSs and Bluetooth for connecting the device to other devices for exchange of the information which can trigger various applications, which can lead this devices attack the targets. If we will use the normal antivirus for malware detection they are totally depending on an updated malware signature database, therefore the static approach like antivirus solution user cannot fully rely on such systems whenever there is attack done with previously un-encountered threat.

A. Problem Specification

Smartphone security is more challenging now a days as in desktop computer systems. As a Practice, Well recognized Security Systems for desktop are considered and evaluation done whether they are applicable to mobile devices or not. In Some of the static systems for desktop like antivirus software are not suitable for Mobile devices as their CPU Consumption is very high also they consume too much memory and the result is rapid draining of the Battery power. If we considered as per the response time of static systems like antivirus it can be between several hours to several days as the process has three steps identifying the new virus, generating a signature, and updating the database, so attackers have high chance of attacking. Many malware instances are also targeting a specific and relatively small features in mobile devices like extracting confidential information and tracking mobile devices location, therefore they are slow movers and difficult to identify. The aim of the pro-

posed system is to design a framework at server side that should detect malwares from the mobile device. Proposed system algorithm extracts, Selects and transforms the features. Information gain algorithm is used for feature selection and SVM algorithm is used for classification and prediction.

4 ALGORITHM OF PROPOSED SYSTEM

In the algorithmic framework use of Support vector machine Algorithm is done for malware detection, First features extraction is done in which features are extracted from mobile device both application and kernel level. For better accuracy and maximum information gain feature selection is done. In which top features are selected from raw features.

There are three Steps of Malware detection framework:

1. Feature Selection
2. Feature Extraction
3. Classification and Prediction

Classification and prediction is a method commonly used in data mining because its efficiency and accuracy. The Aim is to design such framework that can identify the output class of a target based on training dataset as input Values. Each middle node is corresponding to one of the input value; there are edges to various leaf nodes for each of the possible combination of values of that input variable. Each child node represents a value of the target variable which is represented by the edge between the root and the leaf.

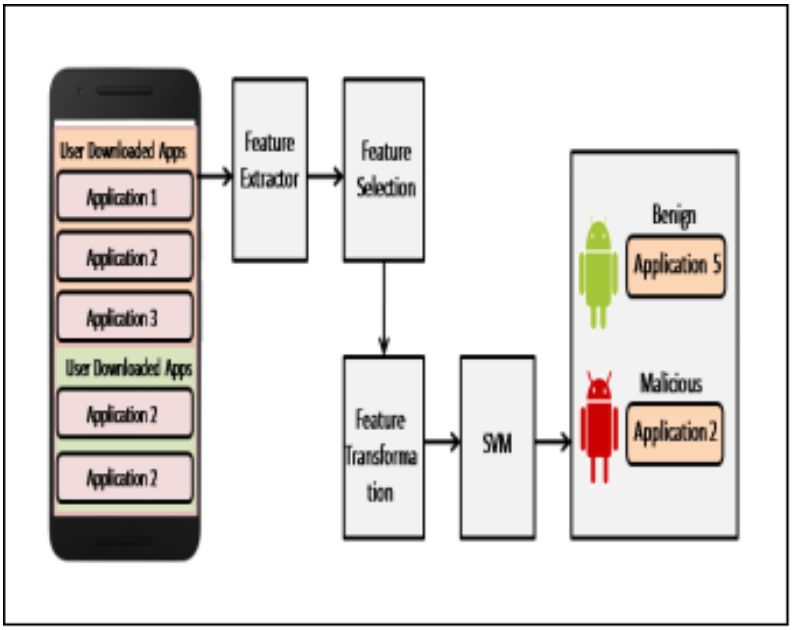

Figure 2 Block Diagram of Proposed System

A. Entropy and Information gain Algorithm

How much information a particular input attributes is giving to determine about the output attribute is Entropy of that attribute. It is also used for a subset of the training data. When we consider information theory, entropy can be a measure of the uncertainty about a Node from where messages are coming. If the receiver is uncertain about a source of messages, the receiver will need more information in order to know sent message. For example, if a Sender is sending always exactly the same message all time, there is no need of any extra information to know about the message. The entropy is always Zero for such a source as there is no uncertainty at all for such messages. In another case, if a sender can send possible messages and for every message is occurring independently of the previous message with same probability distribution, then the receiver's uncertainty is maximized. The amount of information given in bits about the Predicted class is Information gain (IG) about the attribute. When With entropy algorithm features are extracted they can selected with the help of Information gain algorithm.

To improve the accuracy and performance the task of feature selection in these applications is done, but also to cost minimization associated in extracting the features. As number of the features is large in the competition datasets, it can be filtered by algorithm of feature selection in which features are selected first in a single pass and then inductive algorithm is applied. For choice of specific features from applications, as Information gain algorithm has best results amongst all other algorithms.

5 SYSTEM ARCHITECTURE

The proposed system uses Support vector machine technique for realizing a Malware Detection Framework. In the framework, the Feature selector continuously monitors various features and events obtained from the Kernel and then selected features given as input to Support vector machine to classify the application as either benign or malicious.

A. Feature Extraction

In proposed system Data mining Algorithm is used to detect malwares. For testing data set, extracted features from installed applications are considered. Features are divided in two main categories in android operating system.

1. Application Framework - Messaging, Phone Calls, etc.
2. Linux Kernel - Keyboard, Touch Screen, etc.

A. Permissions:

In Android operating system permission is the important aspect as permission decides what activity is to be done by the application in future. These entities include hardware device permissions like camera, GPS as well as operating system permissions like contacts, messages. For example android.permission.INTERNET is a permission which allows the access the Internet and android.permission.READ MASSAGE is a permission which allows the access the user's phone message database. Permissions are divided into two groups: standard and non-standard permissions.

B. Power analyzer:

Monitoring Battery usage is very easy application APIs and It can be used to detect how much amount of power the application

is using. Moreover, Power is strongly related to all the user activities. Since, Battery is a very important resource to keep device active when considered the amount of power used by an application. Rapid draining of power by an application can harm the user as and when there is important work.

C. Traffic Analyzer:

When the permission to use internet is included in manifest file of an application, it can use the network through the mobile device directly. Due to this feature application developers can do misuse by developing malicious application so data traffic monitoring is necessary used by each application. In order to monitor the data traffic use of kernel level APIs can be used and Runtime the traffic can be monitored.

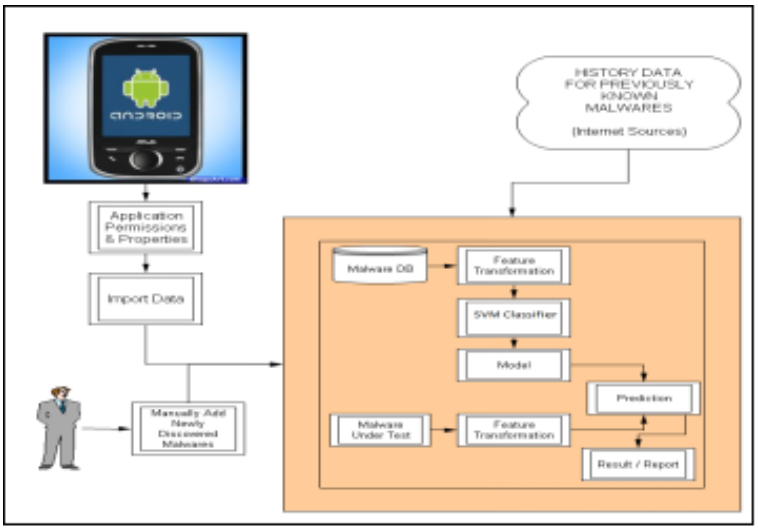

Figure 3 Block Diagram of Proposed System

Feature Selection

In Feature extraction step large number of features is extracted, but amongst them some features are irrelevant or redundant in some cases, if feature section is not done properly it will lead to several problems like increasing model complexity and run-time, reducing generality and misleading the learning algorithm. Effects of these

problems are even more savior when machine Learning methods applied on it. If good feature selection algorithm is applied in a preparatory stage it can help malware detector to classify more efficiently and with a faster detection cycle. Even selected features gives lead to high level of accuracy. Information gain algorithm is most efficient for selection of features from android device, and it gives highest accuracy among other feature selection techniques.

Information Gain (T, X) = Entropy (T) - Entropy (T, X)

Feature Transformation

Feature transformation solves the problem of data compatibility issues between features which are extracted and can be in different formats for generation of log and then the input to the support vector algorithm is provided. Extracted Features can be in different format like Numbers, String, Characters, so cannot be loaded in log file directly. So following types of Feature transformation are done.

1. Converting it to Binary Logic.
2. Converting it to clusters.

For some features binary logic can be used like permissions as they should converted into Yes/NO format. If permission is available for particular activity then it can be converted into 1, and if not then 0. In other case some features are in numbers so for them clustering is to be applied.

6 VI.IMPLEMENTATION OF PROPOSED SYSTEM

Android works on theme of Linux operating system and is an open-source framework which also provides various APIs to most of the hardware components and even for application development .It also allows third-party application developers to design their own applications and they can also be considered. In the implementation of malware detection process various steps are performed like real-time monitoring, feature collection, analysis of various system features,

such as Consumption of CPU, packet analysis through the Wi-Fi, Traffic analysis through internet and battery level.

A. Android Framework

Java programming language is used for applications development in the based on the APIs provided by the Android SDK, but it also allows developers to modify kernel-based functionalities, which is not common for other mobile phone platforms. In Android framework two levels namely User level features and kernel level features are taken in to consideration as a part of feature extraction:

- Permissions
- CPU Usage
- Battery Power
- Traffic Monitoring

B. Threat monitoring unit

The threat monitoring unit communicates with various components of the Android framework, for both Linux kernel as well as the Application Framework layer for collecting various feature metrics, while the Feature Monitoring unit activates the Feature Extractors to collect new feature measurements after specific time interval. Feature monitoring unit also does preprocessing on the some raw features that are collected during Feature Extraction of Kernel level features.

C. SVM methodology

For classification SVM as a machine learning system which will take input from training database for classification of normal and anomaly behavior. Training database is trained through feature extraction from linux powered devices. Various system call generated by processes extracted and input will be given to training database. By learning from training database S.V.M will classify the applications and predict them as anomaly if misbehavior is found. The detector will notify the user and accordingly user can take action on the detected malware.

D. Alert Handler

Alert Handler is an alarm and sends notification as and when required to the user. It activates an action as a result of a dispatched

alert it is a visual alert in the form of message, it also sends notification via SMS or email uninstalling an application, only after the result of classification algorithm alert handler will send notification message.

7 EXPERIMENTS AND RESULTS

Experimental analysis is done in order to evaluate the ability of the proposed malware detection Framework to identify the normal and malicious applications. Experiment contains four sub-experiments, examining the accuracy of the malware detection system for different devices. For each sub experiment random features are extracted from 4 different devices, on which support vector machine algorithm is evaluated. As presented in table 7.1, the features were collected by executing various applications for 8 min each. Experiments are done and system is tested on real devices like Vivo, Samsung, and Sony to understand the better user experience. All the experiments have been performed by considering more than 50 applications. For calculating accuracy of detection algorithm and performance of feature selection algorithm, standard metrics are considered, like True Positive Rate (TPR) and False Positive Rate (FPR). TPR is the proportion of positive instances classified correctly and FPR is the proportion of negative instances misclassified; and Total Accuracy is the proportion of absolutely correctly classified instance; Where,

T1 is number of Normal applications classified correctly.
F1 is the number of malicious applications misclassified.
F2 is the number of Normal applications misclassified.
T2 is the number of malicious applications classified correctly.
TPR = T1 / (T1 + F2) FPR = F1 / (F1 + T2) Total Accuracy = T1 + T1 / T1 + T2 + F1 + F2

TABLE I Experimental results

Exp. No.	Device	Number of Applications	FPR	TPR	AUC	Accuracy
1	Sony Xperia-Z1	78	0.157	0.913	0.928	0.85
2	Samsung Galaxy y	56	0.148	0.862	0.913	0.92
3	Vivo Yl11	77	0.144	0.723	0.851	0.91
4	Samsung Galaxy Note	59	0.132	0.726	0.871	0.83

8 FUTURE ENHANCEMENT

As Smartphone security is one of the emerging field lot of research is going on in the field; as continuous improvement in the android operating systems and new features are coming day by day, attackers are also trying to harm the system so there is enough scope for improvement in security level of Smartphones. Following is list of future work can be carried out based on the proposed work.

1. Other than permissions Wi-Fi packets can be considered as features.
2. Light weight application on phone can be done instead of doing it on server.
3. Light weight application on phone can be done instead of doing it on server.

References

[1] Alzaylaee, Mohammed K., Suleiman Y. Yerima, and SakirSezer. DynaLog: An automated dynamic analysis framework for characterizing Android applications. Cyber Security And Protection Of Digital Services (Cyber Security), IEEE, 2016.

[2] Saracino, Andrea, et al. Madam: Effective and efficient behavior-based android malware detection and prevention. IEEE Transactions on Dependable and Secure Computing, 2016.

[3] Medvet, Eric, and Francesco Mercaldo. Exploring the Usage of Topic Modeling for Android Malware Static Analysis. Availability, Reliability and Security (ARES), IEEE, 2016.

[4] He, Daojing, Sammy Chan, and Mohsen Guizani. Mobile application security: malware threats and defenses. IEEE Wireless Communications, 2015.

[5] Rasthofer, Siegfried, et al. Droidforce: Enforcing complex, data-centric, system-wide policies in android.Availability, Reliability and Security (ARES), IEEE, 2014.

[6] Liang, Shuang, and Xiaojiang Du. Permission-combination-based scheme for android mobile malware detection. Communications (ICC), IEEE, 2014.

[7] Cooper, Vanessa N., Hossain Shahriar, and Hisham M. Haddad. A survey of Android malware characterisitics and mitigation techniques. Information Technology: New Generations (ITNG), IEEE, 2014.

[8] Zheng, Min, Mingshen Sun, and John CS Lui. DroidTrace: a ptrace based Android dynamic analysis system with forward execution capability. Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, 2014.

[9] Li, Li, Alexandre Bartel, Tegawende F. Bissyand. Iccta: Detecting inter-component privacy leaks in android apps. Proceedings of the 37th International Conference on Software Engineering, IEEE Press, 2015.

[10] Dash, Santanu Kumar, et al. Droidscribe: Classifying android malware based on runtime behavior. Security and Privacy Workshops (SPW), IEEE, 2016.

[11] Fratantonio, Yanick, et al. Triggerscope: Towards detecting logic bombs in android applications. Security and Privacy (SP), IEEE, 2016.